A Work Project, presented as part of the requirements for the Award of a Master Degree in Economics from the NOVA School of Business and Economics.

Forecasting Portuguese GDP

A comparison of univariate time series models

José Salvado Garcia, 33088

A Project carried out on the Master in Economics Program, under the supervision of:

João Duarte

03-01-2020

Forecasting Portuguese GDP

A comparison of univariate time series models

Abstract We perform a forecast evaluation exercise, where a broad set of linear univariate models and autoregressive artificial neural networks are compared against a simple linear benchmark when predicting Portuguese real GDP growth. The forecasting exercise is performed in a pseudo-real-time framework, meaning that the specification and estimation of each model are delivered for each quarter of the out-of-sample forecast evaluation interval. The efficacy of the models is tested for diverse conceptions of the loss functions, different evaluation samples, and estimation procedures. The empirical results point to the pre-eminence of artificial neural networks comparatively to linear autoregressions.

Keywords Macroeconomic Forecasting; Time Series Models; Artificial Neural Networks; Structural Breaks

This work used infrastructure and resources funded by Fundação para a Ciência e a Tecnologia (UID/ECO/00124/2013, UID/ECO/00124/2019 and Social Sciences DataLab, Project 22209), POR Lisboa (LISBOA-01-0145-FEDER-007722 and Social Sciences DataLab, Project 22209) and POR Norte (Social Sciences DataLab, Project 22209).

1 Introduction

Macroeconomic projections shape the perspective that agents have respecting economic environment conditions. The expectations formed by these agents concerning present and future economic climate have a fundamental role in the dynamics of aggregate economic variables, such as savings and investment. The developments of monetary and fiscal policy are tightly linked to these aggregate dynamics. It is extensively recognized that monetary and fiscal policy affect both output and inflation with long and variable lags. For this reason, forecasting the direction in which the economy and markets are going is crucial for policymakers. A critical challenge in contemporary economic modelling and forecasting lays on several non-linearities introduced in key macroeconomic time series by structural breaks[1]. Hendry (2000) defends that these developments might represent one of the principal determinants in economic forecast errors.

Over the last years, the Portuguese economy has been confronted by several external and internal shocks that derived on critical structural changes. Gouveia et al. (2018) identify events as the adoption of the euro on 1 January 1999, the European Union enlargements (2004, 2007 & 2013), the Economic and Financial Assistance Programme (agreed in May 2011) as well as the financial and sovereign debt crisis as essential determinants of the Portuguese macroeconomic imbalances that jeopardized economic growth on the past decades[2].

Inspired by the relevance of economic forecasts in policy decision-making, this paper ambition is to analyse whether univariate autoregression models still provide a reliable and valid approach to obtain accurate forecasts for short and medium horizons, or whether considering non-linearities should be brought into the scene when forecasting Portuguese real GDP growth.

The discussion regarding the consideration of non-linearities when forecasting Portuguese economic time series is not something new. For instance, Serra (2018) has already highlighted its meaningfulness when assessing the forecasting performance of the Phillips curves. Rua et al. (2019) focus on the empirical applications of Singular Spectrum Analysis for forecasting

[1] Giraitis et al. (2015) stress the impact that an ineffective consideration of the non-linearities inherent in a time series can have on the forecasting performance.

[2] see Fernandes et al. (2019) for a review on the impact of structural reforms and their repercussion on Portuguese potential GDP and productivity.

quarterly real GDP growth. From the different approaches that can be used to model time-varying non-linear forces, we have identified artificial neural networks as models that have not been sufficiently studied for the Portuguese economy.

The forecasting performance of a comprehensive bundle of autoregressive models and artificial neural networks is analysed by comparing the forecasts (for one, two and four quarter horizons) of the competitor models against a simple autoregressive model. There are considered specifications with different stationarity hypotheses, lag length structures and deterministic components, and the robustness of the results is endorsed by using an ample set of loss functions, evaluation samples and estimation procedures. A pseudo-real-time forecasting exercise is delivered by repeating the specification, estimation and forecasting of every model considered for each quarter in the evaluation period.

We start by looking at the previous work that has been delivered on this field. Starting with relevant international research and going through the different strategies that have been applied for the Portuguese Economy, section 2 provides a literature review on this subject. Section 3 presents the methodology used throughout this work, going through the pseudo-real-time forecasting exercise that was implemented, clarifying the different forecasting models used on the analysis, describing the data used and stressing the forecast performance assessment. The empirical results are presented in section 4 and section 5 compiles the main conclusions retrieved.

2 Literature Review

This section is devoted to reviewing what has been done in the academic field regarding the measurement and comparison of the forecasting performance of different time series models, on an international context and for the Portuguese case. For a complete survey on the different conceptions and applications of time series forecasting see Gooijer and Hyndman (2006).

The evaluation of the forecasting capability of models has been widely studied by economists over the last decades. A considerable body of literature exists, comprehending a far-reaching set of economies and examining different methodologies and applications. One of the most relevant issues studied is whether introducing time-varying and non-linear components can have an advantage over using linear specifications. This approach is justified by the increasingly complex patterns present in economic time series. Shocks derived from political, economic or social changes can impact the dynamics of macroeconomic series[3]. Hence, there is a rising necessity to understand the magnitude and relevance that these structural breaks can imply in economic modelling.

An extensive bundle of methods for assessing the forecasting proficiency of the models have been used by different authors and studies. Early research focused on the role of pooling regressions (Clemen (1989) and Diebold (1989)), where the interest is to understand at which extent can the forecasting ability of a model be enhanced by combining multiple individual forecasts. Forecast encompassing tests have also inspired several studies where the research topic relies on whether the forecast errors of a model can be minimized by considering the forecasts of another model (Chong and Hendry (1986)). For this work project, we have decided to assess the forecasting ability of the competing models based on the relative size of several loss functions associated with forecasting errors. This methodology has already been implemented by several authors, which have compared a large spectrum of models and methods (West (1996), Inoue and Kilian (2006), and Marcellino (2007)).

A usual application consists on comparing simple time series autoregressions with more complex models. Marcellino (2007) studied the performance of several linear, time-varying and nonlinear time series models for forecasting US GDP growth. It was considered a wide set of linear autoregressions, time-varying autoregressions, smooth transition autoregressions and artificial neural networks. None of the sophisticated time-varying and nonlinear models considered was able to beat considerably the linear specifications, which proved to provide an excellent benchmark for other empirical studies and theoretical models.

The benefit from a careful specification of linear time series models was previously noticed by Stock and Watson (1998) when dealing with US macroeconomic time series forecasts. Their work concluded that the forecasts are improved at all horizons when a unit root pretest on the stationarity of the time series is delivered. With respect to nonlinear methods, they found that models with lower parameterization perform best than more tightly parameterized models. For

[3] Marcellino (2007).

short horizons, the use of artificial neural networks delivered lower average losses than the linear models, but the gains were small and only evident after trimming extreme forecasts.

As policymakers have seen their information sets increase considerably during recent years, new methodologies and statistical systems have been developed. Nowadays there are available uncountable time series at a monthly and quarterly basis that can be used to make forecasts that track the fluctuations of the economic and business environment.

Garnitz et al. (2017) proved that GDP forecasts can be enhanced by using survey information. The authors studied the role of a comprehensive and internationally comparable set of indicators (World Economic Survey, conducted by the ifo Institute) to forecast GDP growth in a large set of countries. When the models are fed with information regarding the economic climate and expectations on future economic developments, the forecasts commonly outperform the ones derived from a simple autoregressive model. Furthermore, there are added gains from considering indicators from one of the country's main trading partners. This approach has proved fruitful when nowcasting and forecasting 1 quarter-ahead for the Portuguese case, with gains close to 20% and 15%, respectively.

Moving to more data-intensive methodologies, factor models have tried to extract valuable information from a large dataset. The idea is to summarize the information included in an extensive dataset into a set of common unobserved factors. When using a dataset containing a wide array of economic time series, the factors retrieved from the latter are expected to capture the overall joint movement of the variables, which hopefully will render a close approximation to the economic environment. Relevant studied can be found in Stock and Watson (2002) and Giannone et al. (2008) for the United States and Rünstler et al. (2009) for a European cross-country analysis.

Dias et al. (2015) studied the relative performance of factor models when forecasting GDP growth in Portugal. They found that this approach outperforms significantly the univariate autoregressive model when nowcasting and one-quarter ahead forecasting, rendering gains higher than 50%. Moreover, by using the targeted diffusion index approach developed in Dias et al. (2010) the forecast accuracy is improved.

Rünstler et al. (2009) performed a forecast evaluation exercise for a set of European Economies where large datasets where employed. Under a simulated real-time context, they found that in general factor models outperform bridge equations and quarterly VARs. The models that use monthly data outperformed the ones based on quarterly time series. Nonetheless, the gains when forecasting Portuguese GDP growth are limited.

Moving the spotlight to the use of non-linear conceptions, one of the most studied topics relies on the use of artificial neural networks. An early survey on the application of these models in forecasting can be found in Zhang et al. (1998). Swanson and White (1997) identified that ANNs provide marginal improvements when employed in the forecasting of macroeconomic variables. In the forecasting comparison exercise delivered by Stock and Watson (1998), it is found that these models perform unsatisfactorily when compared with linear specifications. Adebiyi et al. (2014) find that artificial neural networks are able to outperform ARIMA models when modelling stock data. Tkacz (2001) assessed the performance of ANNs when forecasting Canadian GDP growth and found that these models are able to yield statistically lower forecast errors than linear and univariate models.

As a clarification note, it is important to stress the main differences on the approach of the exercise that will be developed throughout this work project and the real-time forecast exercise delivered by central banks and statistical institutions. The main difference that should be pointed is the publication lag that practitioners face in real-time forecasting. Although the univariate time series models considered throughout our analysis provide good benchmarks, real-time practitioners are constantly developing new methods and seeking to improve the performance of their forecasts.

For the particular case of Portugal, Rua and Esteves (2012) present a description of the methodology used by the central bank of the Portuguese republic regarding the short-term forecasting exercise. The authors explain that the standard practice consists in a bottom-up approach, where short-term GDP forecasts are retrieved from forecasts for each of the demand side components in the national accounts (*i.e.* Private and public consumption, Investment, Exports and Imports). The practical application consists of a set of bridge models where the demand side forecast relies on a set of comprehensive economic data and economic indicators that are frequently updated and tuned.

3 Methodology

This section is destined to stress the general principles applied to all models throughout the forecasting exercise.

3.1 Pseudo-real-time Forecast Design

The approach followed throughout this paper follows standard practice and a large series of relevant previous work on forecast evaluation. A pseudo-real-time forecasting exercise consists in repeating the specification, estimation and forecasting of every model considered for each quarter in the evaluation period. This means that all forecasts are recursively estimated using data up to the date of each forecast. For clarification, for the forecasts indexed at quarter Q_t, the estimation sample is $Q_0 - Q_{t-1}$. In the succeeding quarter, the sample is updated to contain the actual GDP growth of quarter Q_t. Our approach is an approximation of the one delivered at real-time, as for the latter practitioners have to deal with publication lags and several reviews on the economic time series used on forecasting.

The first consideration builds upon the selection of a benchmark model. Considering the work of Marcellino (2007) and Stock and Watson (1998) it was decided to select as the most adequate benchmark an autoregression with four lags and a constant. In fact, this specification proved to have a good average performance and its estimation and evaluation are quite straightforward.

The following expression characterizes the generic form of the forecasting model:

$$y\,^i_{t+h} = f\,^i(Z_t; \Theta^i_{ht}) + \epsilon_{t+h}\,i \tag{1}$$

where y_t represents the variable being forecast (in our case, real GDP growth rate), h specifies the forecast horizon, i catalogues the forecasting model (i=1,...,31), Z_t is a vector of predictor variables, Θ_{ht} is a vector of parameters who can be time-varying and ϵ is an error term.

The h-step forecast and the associated forecast error are given by:

$$\hat{y}\,^i_{t+h} = f\,^i(Z_t; \hat{\Theta}\,^i_{ht}) \tag{2}$$

$$e^{\,i}_{\,t+h} = y^{\,i}_{\,t+h} - \hat{y}^{\,i}_{\,t+h} \tag{3}$$

In general, $Z_t = (y_t, ..., y_{t-p}, \Delta y_t, ..., \Delta y_{t-p}, 1, t)$ with t representing the maximum lag lengths. Essentially, the specification of each forecasting model within the methods evaluated will depend both on the assumptions regarding the stationarity of y_t and on the components of Z_t.

Starting by the stationarity assumptions, there are considered specifications where y_t is treated as (possibly trend) stationary or difference stationary. Furthermore, the analysis is complemented with a set of models where the stationarity assumption is decided applying a unit root pre-test. This is a common approach, as literature shows that pretesting routinely might enhance the forecasting performance of the models (Diebold and Kilian (1999)). The stationarity of the series is studied through a set of Augmented Dickey-Fuller tests.

The lag structure is either fixed or chosen by Akaike's Information Criteria (AIC)[4] or Bayesian Information Criteria (BIC)[5] with a maximum number of 6 lags. Preference is given to models with lower information criteria.

As it was noted, the unit-root test, estimation and specification selection for each forecasting model is delivered for each quarter, as to replicate real-time practice.

Regarding the forecasting horizons, the main object of study is focused on short-to-medium range forecasts. Although a more particular scrutiny is delivered for 1 quarter-ahead forecasts, the robustness of the results is complemented by considering horizons of 2 and 4 quarters.

The forecasts derived from the methods are trimmed for the particular case where the absolute forecasted change is extreme. Doing so we are disregarding occasional outliers derived from problems in the specification of the models. When this happens, the forecasted changes are replaced by the average value of the variable.

The available sample period runs from the first quarter of 1995 until the second quarter of 2019. The out-of-sample forecast evaluation interval lies between the first quarter of 2006 and the second quarter of 2019, which corresponds to more than half of the sample period.

[4] $AIC = ln(\hat{\sigma}^2) + \frac{2k}{T}$

[5] $BIC = ln(\hat{\sigma}^2) + \frac{k}{T}ln(T)$, where $\hat{\sigma}^2$ is the residual variance, $k = p + 1$ and T represents the sample size.

Consequently, the first estimation sample (1995q1 to 2005q4) is sufficient to provide a rigorous estimation of the parameters in the forecasting models.

The robustness of the results is achieved by three approaches: First, the use of different loss function to rank the models is useful to study the impact of differentials in the magnitude of the forecast errors. Second, the out-of-sample forecast evaluation span is divided into two subsamples, which are characterized by different economic environments. Last, the recursive approach will be replaced by a rolling mechanism, where a fixed window of 10 years is used as the estimation sample for the quarterly forecasting exercise.

3.2 Forecasting Models

The following section explains the characteristics of the set of forecasting methods that are studied throughout this paper. A detailed description of these can be found in Table 3.

3.2.1 Autoregressive Models

The idea that information in discrete time series processes can be studied having as a basis the concept of stochasticity was popularized by the work of Box and Jenkins (1976).

In fact, the widely common strategy known as the Box-Jenkins approach consists in exploiting the persistence properties of a univariate time series by applying a three-stage scheme for time series identification, estimation and verification.

The specific calibrations within this main forecasting method diverge in the assumptions regarding the stationarity of the y_t variable, the treatment of lag lengths and on the deterministic components to include in the estimation.

Starting by the stationarity assumptions for the autoregressive models, the time series is either specified in levels or in first differences. Additionally, we perform a recursive unit-root pretest on the stationarity of the series. This sums up to three different variations. Second, for the lag length specification, there are considered three different scenarios: It is either fixed at 4 lags or chosen based on AIB or BIC with a maximum of 6 lags. Last, the deterministic component is either composed by a constant or a constant plus a linear trend.

Overall, the derivations inside this category deliver a total of 18 forecasting methods.

3.2.2 No Change

The no change specification consists of a random walk model whose forecast is given by $y_{t+h|t} = y_t$.

3.2.3 Artificial Neural Networks

Artificial Neural Networks models are inspired by the processes intrinsic to the human brain decision-making mechanism. This machine-learning derived method has been used as an approximation of non-linear processes and its forecasting performance has been studied in numerous fields. A broad inspection of the application of such models can be found in the work of Zhang et al. (1998).

The generic forecasting model for this method is given by:

$$y_{t+h} = \beta_0^{'}\,\zeta_t + \sum_{i=1}^{n_1} \gamma_i\, g(\beta_{1i}^{'}\,\zeta_t) + \epsilon_{t+h}$$

The previous expression specifies a feed-forward neural network with one hidden layer. Although the number of hidden layers can be higher, throughout this report there are considered only specifications with one hidden layer, following the suggestion of Kuan and White (1994), which identifies this approach as the most pertinent for economic applications.

The algorithm behind the artificial neural network takes as inputs lags of our target variable to forecast, filters them through a set of neurons inside the hidden layer and determines the combinations that will deliver the output variable - in our case, the forecast. A graphic example of a neural network with three neurons, three lags and one hidden layer can be found in figure 1.

The lag structure for these models is either fixed at three lags or chosen by AIC. These inputs are fed into the n_1 neurons of the hidden layer. The number of neurons varies between one, two and three nodes. y_t is either stationary or differenced.

In order to reach robust forecasts, for each specification within every quarter a total of 1000 networks are fit with different random starting weights. These are then averaged when producing forecasts. Overall, there are studied 12 different conceptions of artificial neural networks.

3.3 Data

The quarterly dataset used on this work consists on the Real Gross Domestic Product (GDP) series for the Portuguese economy. The data is retrieved from the Federal Reserve Bank of St. Louis, for practical convenience. The source of the FRED database is Eurostat, the statistical office of the European Union. The time series is measured in chain-linked volumes (2010) and ranges from the first quarter of 1995 up to the second quarter of 2019. It is both seasonally and calendar adjusted. Figure 2 plots the evolution of GDP between 1995 and 2019. The dynamics of this important macroeconomic variable are considerably stable in the first ten years of the sample. After a negative fluctuation between 2002 and 2003, the dynamics returned to the stable evolution of the preceding years. The developments on the global financial system jeopardized the dynamics of Portuguese GDP after 2008. In 2013 the economy shows evidence of a recovery, although the pace of growth appears to be slower than at the beginning of the sample.

As it has already been unveiled, the target variable throughout the analysis is the quarter-on-quarter real GDP growth rate. When analysing the dynamics of these time series (Figure 3) it can be that it seems stationary. Nonetheless, between 2000 and 2002, and later between 2008 and 2015, the volatility of the series seems higher. Figure 4 plots the differenced series.

These peculiarities are the trigger to the subsample forecast performance evaluation approach, as to make possible a robustness check on the results for periods characterized by different economic conditions.

3.4 Forecast Performance Assessment

When measuring the accuracy of the forecast and comparing the outcomes of different models, there are several approaches that can be implemented. We have decided to perform an out-of-sample model comparison over in-sample approaches, given that when the latter is used the results may be biased towards non-linear models, as their goodness of fit and the capacity of replicating data characteristics may be enlarged by its complex parametrization. Throughout this analysis, we followed the approaches of West (1996), Inoue and Kilian (2006), Marcellino (2007) and Garnitz et al. (2017), which consists in comparing the relative size of a set of loss functions associated with the forecast errors of the different models studied [6].

Three different loss functions are used to rank the models based on their forecasting performance: Mean Squared Error (MSE), Mean Absolute Error (MAE) and Mean Absolute Cubed Error (MACE). The robustness of the analysis is enhanced when including three different loss functions, as the weight associated with the magnitude of the errors is different between them.

The main argument of such loss functions is the forecasting error ($e\,^{i}_{t+h}$) produced by the models. It is given by the difference between the observed value and the forecast (equation 3).

In order to rank the forecasting accuracy of the competing models, it is computed the relative forecast error[7] for each of the loss functions in analysis.

Whenever the relative forecast error is smaller than one, the competing model i has, on average, a higher forecast accuracy compared with the benchmark model. On the other hand, if the ratio is higher than one the model is not able to overcome the accuracy of the benchmark.

3.4.1 Mean Squared Error (MSE)

The Mean Squared Error is one of the most common measures of forecast accuracy used in general practice and it calculates the average of the squares of the forecasting error. This measure is a function of both the variance and the bias of the errors. The MSE is defined as

$$MSE\,^{j}_{i,h} = \frac{1}{N}\sum_{n=1}^{N}(e\,^{i}_{t+h})^2 \tag{4}$$

3.4.2 Mean Absolute Error (MAE)

The Mean Absolute Error (MAE) is similar to the MSE as it also measures the magnitude of the errors without considering their direction, but in the previous larger weights are assigned to larger errors.

[6] Other relevant studies propose alternative comparison methods, as forecast encompassing tests (Chong and Hendry (1986)).

[7] $Relative\ Forecasting\ Error\,^{i} = \frac{Loss\ Funtion\,^{i}}{Loss\ Funtion\,^{Benchmark}}$

It is given by:

$$MAE\,^{j}_{i,h} = \frac{1}{N}\sum_{n=1}^{N} |e\,^{i}_{t+h}| \tag{5}$$

3.4.3 Mean Absolute Cubed Error (MACE)

Last, the Mean Absolute Cubed Error provides a larger weight to larger errors than the two previous loss functions, and it is computed as follows:

$$MACE\,^{j}_{i,h} = \frac{1}{N}\sum_{n=1}^{N} |e\,^{i}_{t+h}|^{3} \tag{6}$$

4 Results

The coming section is destined for the discussion of the main implications retrieved from the forecasting evaluation exercise. The results of the out-of-sample forecast assessment for the linear and non-linear methods are presented in Tables 1 and 2, respectively.

Overall, a set of nineteen linear specifications and twelve artificial neural networks are compared against a baseline model (an autoregressive model in levels with 4 lags and a constant term). The comparison exercise is delivered based on each model's forecast for the succeeding quarters (one, two and four-quarter ahead forecasts).

For the recursive estimation, results are reported for all forecast horizons and loss functions. Additionally, a robustness check is delivered by splitting the out-of-sample evaluation period into two subsamples, characterized by contrasting economic conditions. Also, results are summarized for the case where the models are estimated using a rolling window of 10 years, which was recursively updated. Pesaran and Timmermann (1995) commit to a rolling estimation sample as a way of minimizing the effects of structural changes.

Table 4 exhibits the results of the Augmented Dickey-Fuller tests for the stationarity assumptions respecting Portuguese GDP Growth. The results are presented for tests which incorporate a constant and a constant plus a trend in the general regression equation, up to 6 lags. Table 5 presents the resulting optimal lag structures of the linear models, selected both by AIC and BIC for the different stationarity assumptions. The combination of the outputs in the aforementioned tables lead to the selection of the pre-tested linear models.

The best performing models are showcased in Figure 5, for all forecasting horizons.

4.1 Linear Models

On a first stance, the results for the linear models are reported. Starting by comparing the alternative specifications regarding their 1-quarter ahead MSE relative to the benchmark (column 2 of Table 1), the results point to a relative gain when adding a linear trend in the estimation of models that handle real GDP growth as a stationary process. These gains are maximized when selecting the lag structure of the model based on information criteria. In general, models where GDP growth is treated as non-stationary perform worse than models specified in levels. The best[8] performing model under this setting is an autoregression with constant, trend and lag structure determined by the BIC criteria (ARFT0b), whose relative MSE is 0.94. Notwithstanding, the average gain is not significant.

When we consider different loss functions (MAE and MACE, columns 3 and 4, respectively) the previous results do not hold, as the ARFT0b does not longer beat the benchmark. There is a slight benefit from modelling in first differences, with the best competitor being the ARFC1b. In effect, its relative MAE is 0.98 and the relative MACE is 0.97.

Overall, although some models present gains when compared with the benchmark model, the magnitude of the gains is not astonishing, as choosing a different specification never renders gains higher than 10%.

When considering larger forecast horizons, advances are found for models specified in levels, assuming stationarity on the time series and adding a linear trend as a deterministic component. For the 2-quarter ahead forecasts, the model that presents a lower relative MSE is an ART0a (0.81). The best performer on a 4-quarter ahead forecasting exercise is an ARFT04 (0.85). These conclusions are consistent when considering different loss functions, although the relative gain is lower.

The same forecast performance assessment was delivered for two different subsamples. The

[8]When a set of models have an identical relative loss function, priority will be given to the most simple specification.

main forecasting period was split into 2006q1-2012q4 and 2013q1-2019q2. The first is described by an unstable situation on the Portuguese economy, when GDP growth was impaired by the Global Financial Crisis and subsequent debt crisis. The second subsample represents a period of recovery, with lower volatility in Portuguese GDP growth dynamics.

Starting by the 2006q1 to 2012q4 subsample, a considerable number of specifications are able to beat the benchmark. The model with the lowest relative MSE is again the ARFT0b, with a relative MSE of 0.84. This performance was also reached by the ARFCP4 specification. When looking at the results for the second sub-sample (2013q1-2019q2), the best model is the ARFC0b specification, with a relative MSE of 0.89. For this subsample, there are fewer models beating the benchmark.

The results of the forecast evaluation under a rolling estimation for a 1-quarter ahead forecasts are provided in column 13 of Table 1. The models with the lowest relative MSE are the ARFC0b and ARFT0b specifications (0.91). When compared with the results under the recursive estimation, more models outperform the benchmark and the average gain is higher. This supports the idea that estimating under a rolling sample reduces the impact of structural breaks.

In general, we find evidence of added value when deciding the lag specification based on information criteria, for models that include a linear trend in the estimation. The improvements in the forecasting performance increase when considering more distant time spans. However, the magnitude of the average forecasting gains is never higher than 20%. These results are in line with the findings of Marcellino (2007) for US GDP growth and Rünstler et al. (2009) for the Portuguese case.

4.2 Artificial Neural Networks

The results of the out-of-sample evaluation of the artificial neural network specifications are presented in Table 2, in the same logic as the linear models.

Starting by the 1-quarter ahead forecasts (column 2 of Table 2), the artificial neural networks specified in levels and with fixed lag structure are able to consistently overcome the benchmark. The best model within this framework is the specification with only one hidden layer (NNAR013), whose relative MSE is 0.84. The results for this forecasting horizon are robust when considering different loss functions. In particular, the NNAR013 model presents a higher relative MAE (0.89) but lower relative MACE (0.79).

When considering larger forecast horizons, the relative gains of using the artificial neural networks increase. For the 2-quarter ahead forecasts, the model that presents a lower relative MSE is again the NNAR03 (0.80), although all models specified in levels and with fixed lag structures beat the linear benchmark. The forecasting relative gains under a recursive estimation are maximized when considering a 4-quarter ahead forecasting exercise, where the NNAR013 yields a relative MSE of 0.76. These results are robust when examining different loss functions.

The recursive split analysis delivers interesting results, as almost all artificial neural networks are able to beat the benchmark in the first subsample. Recall that this subsample is characterized by structural changes. The best performer is the NNAR013, whose relative MSE is 0.78. On the other hand, none of the specifications can overcome the benchmark on the second subsample, characterized by a recovery on the dynamics of the Portuguese economy.

The results of the forecast evaluation under a rolling estimation for a 1-quarter ahead forecasts are provided in column 13 of Table 2. The model with the lowest relative MSE is the NNAR013 (0.75) and most of the linear conceptions of the artificial neural networks are able to beat the benchmark.

Overall, we find evidence of forecasting gains when using artificial neural networks for modelling Portuguese real GDP growth dynamics. These models can considerably outperform the linear autoregressions studied in this paper. An important conclusion is the pertinence of such models under periods typified by structural changes. The improvements in the forecasting performance increase when considering more distant forecasts and can deliver improvements up to 30%. Notwithstanding, the increase in the forecasting performance found by other authors using alternative specifications to model non-linearities is higher. For instance, the factor models used by Dias et al. (2015) yield wins up to 70%.

5 Conclusion

Throughout this paper, it was delivered an inspection on the advantage of using refined autoregressions and non-linear time series models for modelling Portuguese GDP growth.

Altogether, the empirical results support that more sophisticated forecasting models as the artificial neural networks have wider gains when compared with a comprehensive bundle of linear autoregressions. The artificial neural networks are especially relevant under periods typified by structural changes. Our results reveal that employing a rolling estimation provides a powerful strategy to overcome issued derived from structural changes.

Consequently, modern economic modelling and forecasting must consider the presence of non-linearities in macroeconomic time series, in order to minimize the errors in the forecasts produced by the emergence of structural changes.

References

Adebiyi, A. A., Adewumi, A. O., and Ayo, C. K. (2014). Comparison of arima and artificial neural networks models for stock price prediction. *Journal of Applied Mathematics*, 2014.

Box, G. and Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control*. Holden-Day.

Chong, Y. Y. and Hendry, D. F. (1986). Econometric evaluation of linear macro-economic models. *The Review of Economic Studies*, 53(4):671–690.

Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting*, 5(4):559 – 583.

Dias, F., Pinheiro, M., and Rua, A. (2010). Forecasting using targeted diffusion indexes. *Journal of Forecasting*, 29(3):341–352.

Dias, F., Pinheiro, M., and Rua, A. (2015). Forecasting portuguese GDP with factor models: Pre-and post-crisis evidence. *Economic Modelling*, 44:266–272.

Diebold, F. X. (1989). Forecast combination and encompassing: Reconciling two divergent literatures. *International Journal of Forecasting*, 5(4):589 – 592.

Diebold, F. X. and Kilian, L. (1999). Unit root tests are useful for selecting forecasting models. Working Paper 6928, National Bureau of Economic Research.

Fernandes, A., Simões, M., and Pereira, J. A. (2019). Alterações estruturais da economia portuguesa, produto potencial e produtividade. *Gabinete de Planeamento, Estratégia, Avaliação e Relações Internacionais*.

Garnitz, J., Lehmann, R., and Wohlrabe, K. (2017). Forecasting gdp all over the world: Evidence from comprehensive survey data.

Giannone, D., Reichlin, L., and Small, D. (2008). Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics*, 55(4):665–676.

Giraitis, L., Kapetanios, G., Mansur, M., and Price, S. (2015). *Forecasting Under Structural Change*, pages 401–419. Springer International Publishing, Cham.

Gooijer, J. G. D. and Hyndman, R. J. (2006). 25 years of time series forecasting. *International Journal of Forecasting*, 22(3):443 – 473. Twenty five years of forecasting.

Gouveia, A., Coelho, M., Cruz, D., , Fernandes, A., Osterhold, C., Pereira, E., Leal, F., Pereira, J., Simões, M., Rodrigues, R., and Martins, T. (2018). The portuguese economy: Short essays on structural changes. *Gabinete de Planeamento, Estratégia, Avaliação e Relações Internacionais*.

Hendry, D. F. (2000). On detectable and non-detectable structural change. *Structural change and economic dynamics*, 11(1-2):45–65.

Inoue, A. and Kilian, L. (2006). On the selection of forecasting models. *Journal of Econometrics*, 130(2):273–306.

Kuan, C.-M. and White, H. (1994). Artificial neural networks: an econometric perspective. *Econometric Reviews*, 13(1):1–91.

Marcellino, M. (2007). A comparison of time series models for forecasting GDP growth and inflation. *Bocconi University, Italia*.

Pesaran, M. H. and Timmermann, A. (1995). Predictability of stock returns: Robustness and economic significance. *The Journal of Finance*, 50(4):1201–1228.

Rua, A. and Esteves, P. (2012). Short-term forecasting for the portuguese economy: a methodological overview. *Economic Bulletin and Financial Stability Report Articles and Banco de Portugal Economic Studies*.

Rua, A., Hassani, H., Silva, E. S., and Thomakos, D. (2019). Monthly Forecasting of GDP with Mixed Frequency Multivariate Singular Spectrum Analysis. Technical report.

Rünstler, G., Barhoumi, K., Benk, S., Cristadoro, R., Den Reijer, A., Jakaitiene, A., Jelonek, P., Rua, A., Ruth, K., and Van Nieuwenhuyze, C. (2009). Short-term forecasting of gdp using large datasets: a pseudo real-time forecast evaluation exercise. *Journal of forecasting*, 28(7):595–611.

Serra, S. (2018). Is the Phillips curve dead? - results for Portugal. *Economic Bulletin and Financial Stability Report Articles and Banco de Portugal Economic Studies*.

Stock, J. H. and Watson, M. W. (1998). A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series. *National Bureau of Economic Research*.

Stock, J. H. and Watson, M. W. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American statistical association*, 97(460):1167–1179.

Swanson, N. R. and White, H. (1997). A model selection approach to real-time macroeconomic forecasting using linear models and artificial neural networks. *Review of Economics and Statistics*, 79(4):540–550.

Tkacz, G. (2001). Neural network forecasting of canadian GDP growth. *International Journal of Forecasting*, 17(1):57–69.

West, K. D. (1996). Asymptotic inference about predictive ability. *Econometrica: Journal of the Econometric Society*, pages 1067–1084.

Zhang, G., Patuwo, B. E., and Hu, M. Y. (1998). Forecasting with artificial neural networks: The state of the art. *International journal of forecasting*, 14(1):35–62.

Annexes

	Recursive $h=1$			Recursive $h=2$			Recursive $h=4$			Recursive Split $h=1$		Rolling $h=1$
Model	MSE	MAE	MACE	MSE	MAE	MACE	MSE	MAE	MACE	MSE	MSE	MSE
ARFC04	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
ARFT04	0.96	1.01	1.03	0.88	0.99	0.98	**0.85**	**0.96**	**0.93**	0.87	1.21	0.95
ARFC14	1.12	1.04	1.07	1.17	1.06	1.12	1.26	1.08	1.16	1.02	1.44	1.11
ARFT14	1.13	1.04	1.08	2.13	1.41	1.98	5.89	2.45	6.02	1.03	1.45	1.04
ARFC0a	1.00	1.01	1.03	1.06	1.03	1.06	1.04	1.02	1.04	1.02	0.93	0.94
ARFC0b	1.00	1.02	1.04	1.06	1.01	1.03	0.98	0.98	0.96	1.03	**0.89**	**0.91**
ARFC1a	1.10	1.05	1.10	1.14	1.08	1.16	1.31	1.20	1.45	1.01	1.37	1.09
ARFC1b	1.05	**0.98**	**0.97**	1.07	0.99	0.98	1.31	1.12	1.26	0.94	1.39	1.03
ARFT0a	0.95	1.02	1.04	**0.81**	**0.96**	**0.93**	0.89	0.98	0.97	0.85	1.26	0.92
ARFT0b	**0.94**	1.00	1.01	0.85	0.97	0.94	0.93	0.98	0.97	**0.84**	1.19	**0.91**
ARFT1a	1.13	1.06	1.13	2.36	1.54	2.37	6.82	2.75	7.55	1.05	1.40	1.13
ARFT1b	1.06	0.99	**0.97**	1.89	1.33	1.78	5.94	2.53	6.38	0.95	1.40	1.04
ARFCP4	1.06	1.05	1.11	1.26	1.05	1.11	1.36	1.17	1.36	**0.84**	1.75	0.95
ARFTP4	1.10	1.03	1.06	1.15	1.03	1.07	1.25	1.06	1.13	1.02	1.34	1.11
ARFCPa	1.03	1.01	1.03	1.66	1.17	1.37	4.39	1.82	3.31	1.05	0.99	1.02
ARFTPa	1.07	1.02	1.03	1.10	1.03	1.07	1.32	1.17	1.36	1.02	1.24	1.01
ARFCPb	1.02	1.00	1.00	1.62	1.16	1.35	4.32	1.83	3.36	1.04	0.98	1.01
ARFTPb	1.10	1.02	1.04	1.19	1.06	1.12	1.53	1.25	1.57	1.03	1.33	0.99
No Change	1.01	1.01	1.02	1.50	1.13	1.27	4.54	1.89	3.58	0.98	1.10	0.98

Note: Shaded - models that outperform the benchmark; Bold - Best model performing

Table 1: Linear models out-of-sample evaluation

	Recursive $h=1$			Recursive $h=2$			Recursive $h=4$			Recursive Split $h=1$		Rolling $h=1$
Model	MSE	MAE	MACE	MSE	MAE	MACE	MSE	MAE	MACE	MSE	MSE	MSE
ARFC04	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
NNAR013	**0.84**	**0.89**	**0.79**	**0.80**	**0.89**	**0.79**	**0.76**	**0.85**	**0.72**	**0.78**	1.04	**0.75**
NNAR023	0.95	0.92	0.85	0.88	0.90	0.82	0.91	0.95	0.90	0.86	1.21	0.84
NNAR033	0.99	0.94	0.88	0.86	0.91	0.83	1.06	1.05	1.10	0.91	1.23	0.88
NNAR01a	1.02	1.03	1.06	0.88	0.97	0.95	0.88	1.02	1.03	0.93	1.32	0.91
NNAR02a	1.01	1.03	1.06	0.93	1.01	1.02	1.08	1.11	1.24	0.91	1.31	0.90
NNAR03a	1.06	1.06	1.13	0.96	1.03	1.06	1.23	1.19	1.41	0.94	1.44	0.94
NNAR113	1.23	1.05	1.09	1.12	1.01	1.03	1.36	1.14	1.30	1.15	1.47	0.93
NNAR123	1.20	1.04	1.08	1.20	1.05	1.10	1.55	1.24	1.53	1.11	1.48	1.12
NNAR133	1.01	1.03	1.05	1.18	1.03	1.06	1.57	1.24	1.53	0.83	1.53	1.21
NNAR11a	1.15	1.05	1.10	1.22	1.08	1.16	1.43	1.31	1.71	0.90	1.91	1.02
NNAR12a	1.21	1.08	1.16	2.06	1.23	1.50	1.58	1.32	1.75	0.93	2.06	3.24
NNAR13a	1.21	1.08	1.16	4.16	1.40	1.97	2.14	1.45	2.09	0.93	2.06	4.35

Note: Shaded - models that outperform the benchmark; Bold - Best model performing

Table 2: Artificial neural networks out-of-sample evaluation

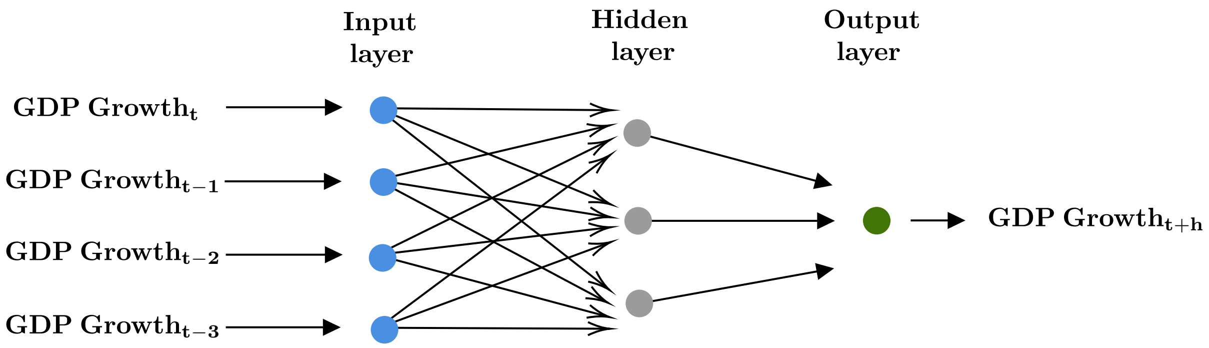

Figure 1: A three lagged autoregressive neural network with 1 hidden layer with 3 nodes

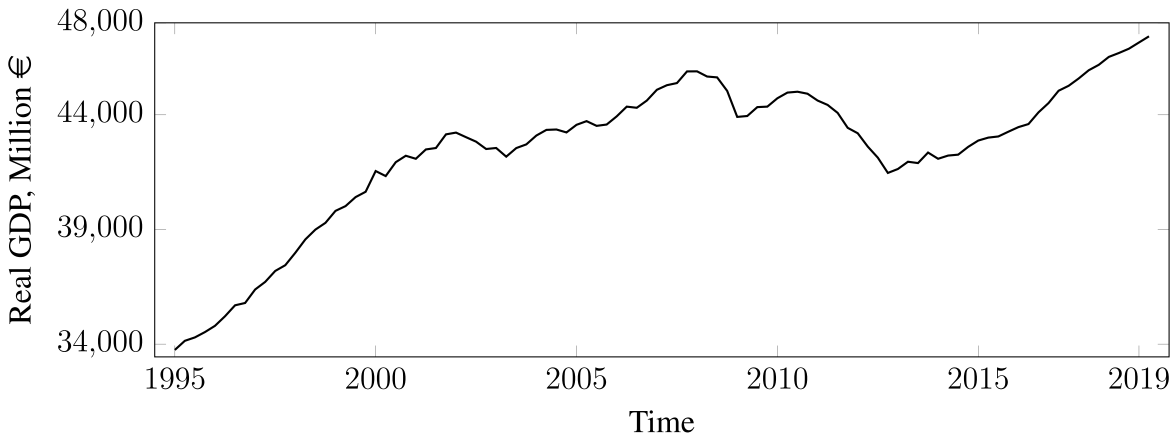

Figure 2: Quarterly real Gross Domestic Product for Portugal from 1995Q1 to 2019Q2

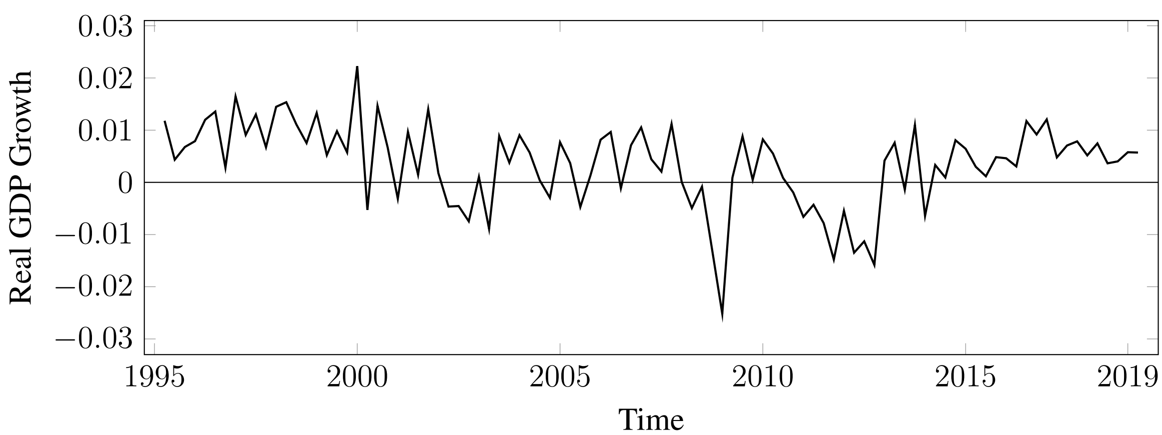

Figure 3: Quarterly real Gross Domestic Product growth for Portugal from 1995Q2 to 2019Q2

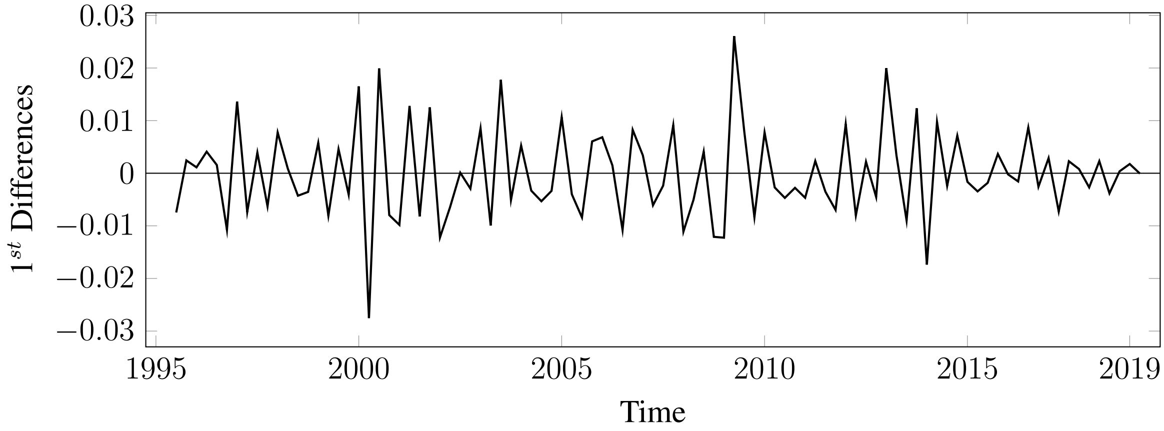

Figure 4: Differenced quarterly real Gross Domestic Product growth for Portugal from 1995Q3 to 2019Q2

Code	Description
Autoregressive Models	
AR(*d, u, p*)	Autoregressions
	d = deterministic components included
	= C (constant only) or T (contant and linear trend)
	u = stationarity assumptions
	= 0 (levels), 1 (differences) or P (Pre-tested for a unit root)
	p = number of lags
	= 4 lags , A (AIC, $0 \leq p \leq 6$) or B (BIC, $0 \leq p \leq 6$)
No Change	
NOCHANGE	$y_{t+h\|t} = y_t$
Neural Networks	
NN(*u, n, p*)	
	u = stationarity assumptions
	= 0 (levels), 1 (differences)
	n = number of hidden networks
	= 1, 2 or 3
	p = number of lags
	= 3 lags or A (AIC, $0 \leq p \leq 6$)

Table 3: Summary of forecasting methods

	ADF with drift							ADF with drift and trend						
Lag	0	1	2	3	4	5	6	0	1	2	3	4	5	6
2006q1	0	0	1	1	1	1	1	0	0	1	1	1	1	1
2006q2	0	0	1	1	1	1	1	0	0	1	1	1	1	1
2006q3	0	0	1	1	1	1	1	0	0	1	1	1	1	1
2006q4	0	0	1	1	1	1	1	0	0	1	1	1	1	1
2007q1	0	0	1	1	1	1	1	0	0	1	1	1	1	1
2007q2	0	0	1	1	1	1	1	0	0	1	1	1	1	1
2007q3	0	0	1	1	1	1	1	0	0	1	1	1	1	1
2007q4	0	0	1	1	1	1	1	0	0	1	1	1	1	1
2008q1	0	0	1	1	1	1	1	0	0	1	1	1	1	1
2008q2	0	0	1	1	1	1	1	0	0	1	1	1	1	1
2008q3	0	0	1	1	1	1	1	0	0	1	1	1	1	1
2008q4	0	1	1	1	1	1	1	0	0	1	1	1	1	1
2009q1	0	1	1	1	1	1	1	0	1	1	1	1	1	1
2009q2	0	1	1	1	1	1	1	0	0	1	1	1	1	1
2009q3	0	0	1	1	1	1	1	0	0	0	1	1	0	1
2009q4	0	0	1	1	1	1	1	0	0	0	0	0	0	1
2010q1	0	0	0	1	1	0	1	0	0	0	0	0	0	1
2010q2	0	0	1	1	1	1	1	0	0	0	0	0	0	1
2010q3	0	0	0	1	1	1	1	0	0	0	0	0	0	1
2010q4	0	0	0	1	1	1	1	0	0	0	0	0	0	1
2011q1	0	0	1	1	1	1	1	0	0	0	0	0	0	1
2011q2	0	0	1	1	1	1	1	0	0	0	0	0	0	1
2011q3	0	0	1	1	1	1	1	0	0	0	0	0	0	1
2011q4	0	0	1	1	1	1	1	0	0	0	0	0	0	1
2012q1	0	0	1	1	1	1	1	0	0	0	0	0	0	0
2012q2	0	0	1	1	1	1	1	0	0	0	0	0	0	1
2012q3	0	0	1	1	1	1	1	0	0	0	0	0	0	1
2012q4	0	1	1	1	1	1	1	0	0	0	0	0	0	1
2013q1	0	0	1	1	1	1	1	0	0	0	0	0	0	0
2013q2	0	0	0	1	0	0	1	0	0	0	0	0	0	0
2013q3	0	0	0	1	0	0	1	0	0	0	0	0	0	0
2013q4	0	0	0	1	0	0	1	0	0	0	0	0	0	0
2014q1	0	0	0	1	0	0	1	0	0	0	0	0	0	0
2014q2	0	0	0	1	0	0	1	0	0	0	0	0	0	0
2014q3	0	0	0	0	0	0	1	0	0	0	0	0	0	0
2014q4	0	0	0	0	0	0	1	0	0	0	0	0	0	1
2015q1	0	0	0	0	0	0	1	0	0	0	0	0	0	1
2015q2	0	0	0	0	0	0	1	0	0	0	0	0	0	1
2015q3	0	0	0	0	0	0	1	0	0	0	0	0	0	1
2015q4	0	0	0	0	0	0	1	0	0	0	0	0	0	1
2016q1	0	0	0	0	0	0	1	0	0	0	0	0	0	1
2016q2	0	0	0	0	0	0	1	0	0	0	1	1	0	1
2016q3	0	0	0	0	0	0	1	0	0	0	1	1	0	1
2016q4	0	0	0	0	0	0	1	0	0	1	1	1	1	1
2017q1	0	0	0	0	0	0	1	0	0	1	1	1	1	1
2017q2	0	0	0	0	0	0	1	0	0	1	1	1	1	1
2017q3	0	0	0	0	0	0	1	0	0	1	1	1	1	1
2017q4	0	0	0	0	0	0	1	0	0	1	1	1	1	1
2018q1	0	0	0	0	0	0	1	0	0	1	1	1	1	1
2018q2	0	0	0	0	0	0	1	0	0	1	1	1	1	1
2018q3	0	0	0	0	0	0	1	0	0	1	1	1	1	1
2018q4	0	0	0	0	0	0	1	0	0	1	1	1	1	1
2019q1	0	0	0	0	0	0	1	0	0	1	1	1	1	1
2019q2	0	0	0	0	0	0	1	0	0	1	1	1	1	1

Table 4: Results of the Augmented Dickie-Fuller test on the stationarity of Portuguese real GDP growth

	ARFC0		ARFC1		ARFT0		ARFT1	
	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
2006q1	2	2	2	2	3	2	2	2
2006q2	2	2	2	1	2	2	2	1
2006q3	2	2	2	2	3	2	2	2
2006q4	2	2	2	2	3	2	2	2
2007q1	2	2	2	2	3	2	2	2
2007q2	2	2	2	2	3	2	2	2
2007q3	2	2	3	2	3	2	6	2
2007q4	2	2	3	2	3	2	3	2
2008q1	3	2	3	2	3	2	3	2
2008q2	3	2	3	2	3	3	6	2
2008q3	2	2	2	2	3	2	6	2
2008q4	3	2	2	2	3	2	2	2
2009q1	3	2	2	2	3	2	2	2
2009q2	3	2	2	2	3	3	2	2
2009q3	2	1	6	2	3	2	6	2
2009q4	2	1	6	2	3	2	6	2
2010q1	2	1	6	2	3	2	6	2
2010q2	2	1	6	2	3	2	6	2
2010q3	2	1	6	2	3	2	6	2
2010q4	2	1	6	2	3	2	6	2
2011q1	2	1	6	2	3	2	6	2
2011q2	2	1	6	2	3	2	6	2
2011q3	2	2	6	2	3	2	6	2
2011q4	2	2	6	2	3	2	6	2
2012q1	2	2	6	2	3	2	6	2
2012q2	3	2	6	2	3	2	6	2
2012q3	3	2	6	2	3	2	6	2
2012q4	3	2	6	2	3	2	6	2
2013q1	2	2	6	2	3	2	6	2
2013q2	2	2	6	2	2	2	6	2
2013q3	2	2	6	2	3	2	6	2
2013q4	2	2	6	2	3	2	6	2
2014q1	2	2	6	2	3	2	6	2
2014q2	2	2	6	2	3	2	6	2
2014q3	2	2	6	2	3	2	6	2
2014q4	2	2	6	2	2	2	6	2
2015q1	2	2	6	2	2	2	6	2
2015q2	2	2	6	2	2	2	6	2
2015q3	2	2	6	2	2	2	6	2
2015q4	2	2	6	2	2	2	6	2
2016q1	2	2	6	2	2	2	6	2
2016q2	2	2	6	2	2	2	6	2
2016q3	2	2	6	2	3	2	6	2
2016q4	2	2	6	2	3	2	6	2
2017q1	2	2	6	2	2	2	6	2
2017q2	2	2	6	2	2	2	6	2
2017q3	2	2	6	2	2	2	6	2
2017q4	2	2	6	2	3	2	6	2
2018q1	2	2	6	2	3	2	6	2
2018q2	2	2	6	2	3	2	6	2
2018q3	2	2	6	2	3	2	6	2
2018q4	2	2	6	2	3	2	6	2
2019q1	2	2	6	2	3	2	2	2
2019q2	2	2	6	2	3	2	6	2

Table 5: Determining the optimal lag structure through Akaike's and Bayesian Information Criteria

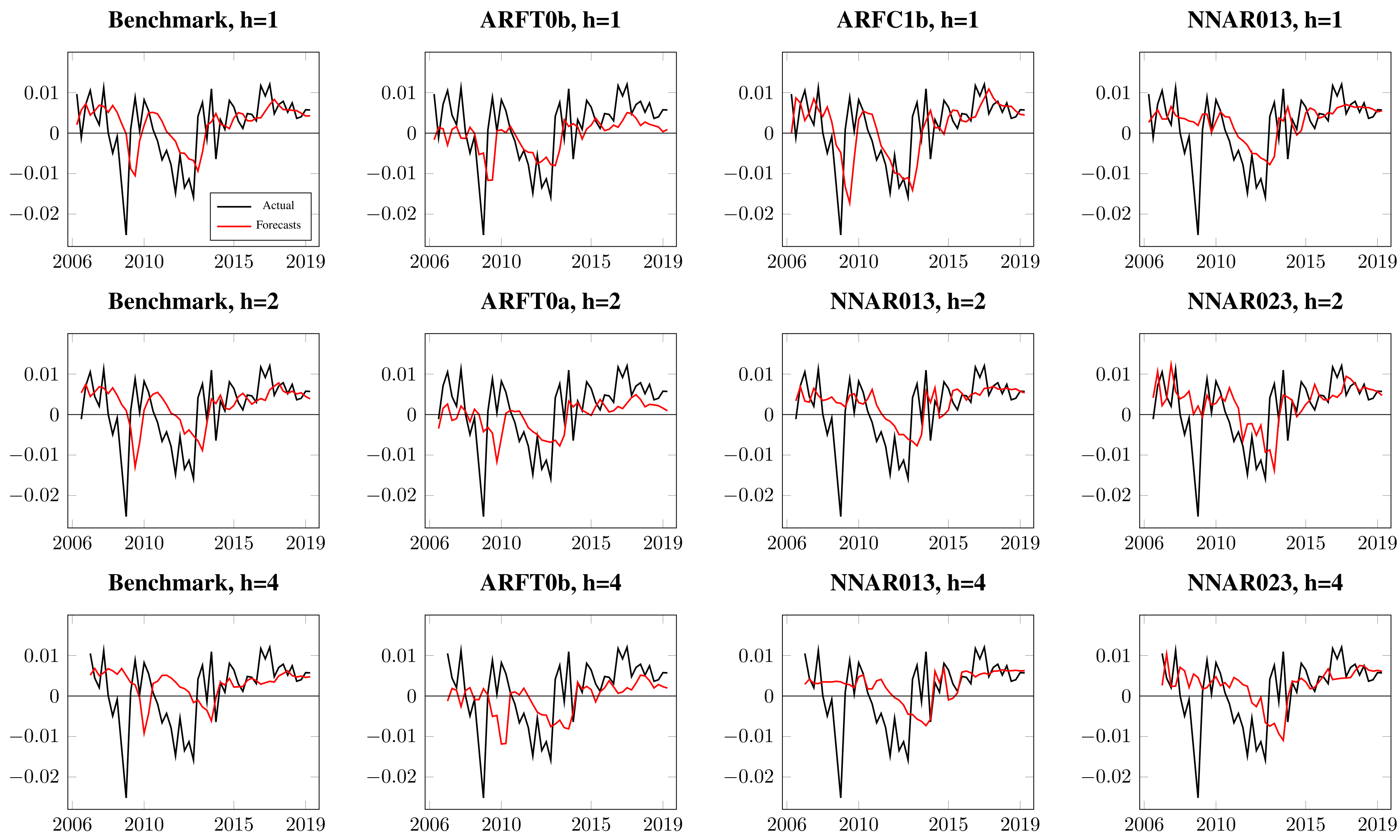

Figure 5: Plotting the forecasted values of Portuguese real GDP growth for the benchmark and a selection of best performing models for each quarter in analysis